Data Science and Femtography

Peter Alonzi, UVA, Data Science Institute

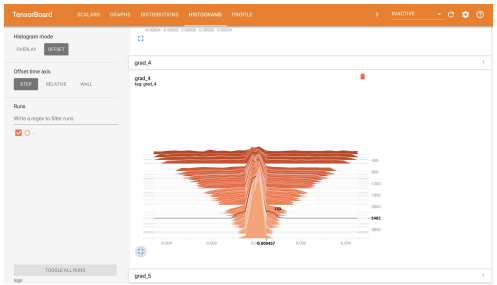
CNF2019 Symposium, SURA Headquarters

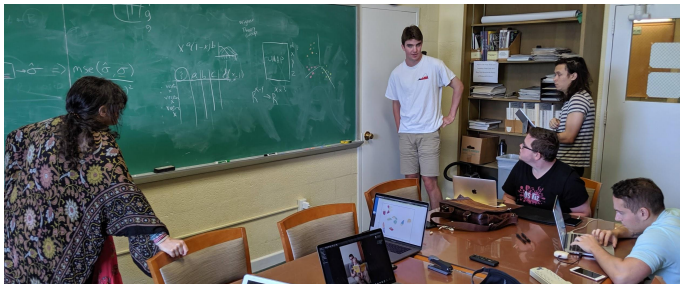
CNF19-14: Sponsored by Southeastern Universities Research Association, Inc. Grant Agreement No. C2019-FEMT-002-04

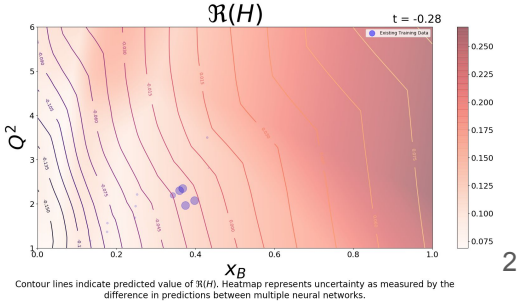
August 12 and 13, 2019, Washington DC

Summer Institute for Wigner Imaging and Femtography

1. Data Science enhances Physics
2. UVA Data Science
3. Interdisciplinary development
4. Machine Learning applications







Contour lines indicate predicted value of $\Re(H)$. Heatmap represents uncertainty as measured by the difference in predictions between multiple neural networks.

Data Science and Physics

- Data science craves data → scaleable to petabytes and beyond
- Build models and test hypotheses → scientific method
- Machine learning uses probabilities → natural fit for quantum systems
- ...

$$\langle DataScience | Physics \rangle \approx 1$$

Data Science at UVA

“An interdisciplinary field that uses scientific methods to extract knowledge and insights from data.”

- School of Data Science plans announced
- Mandate for Interdisciplinary efforts
- Today: ~50 MSDS students per cohort
- Tomorrow: PhD students

Phil Bourne, Director, Data Science Institute

Data Science at UVA

“An interdisciplinary field that uses scientific methods to extract knowledge and insights from data.”

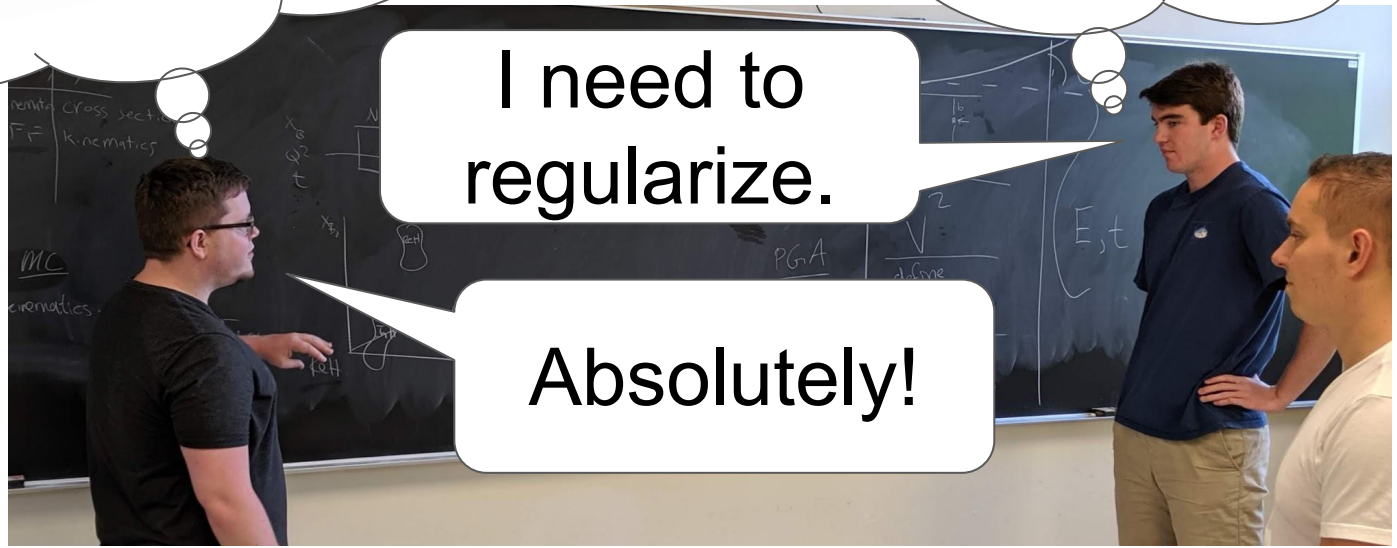
- Data management
- Communication
- Visualization (don’t miss VR demo!!!)
- Machine learning

Phil Bourne, Director, Data Science Institute

Interdisciplinarity is hard - e.g. Regularization

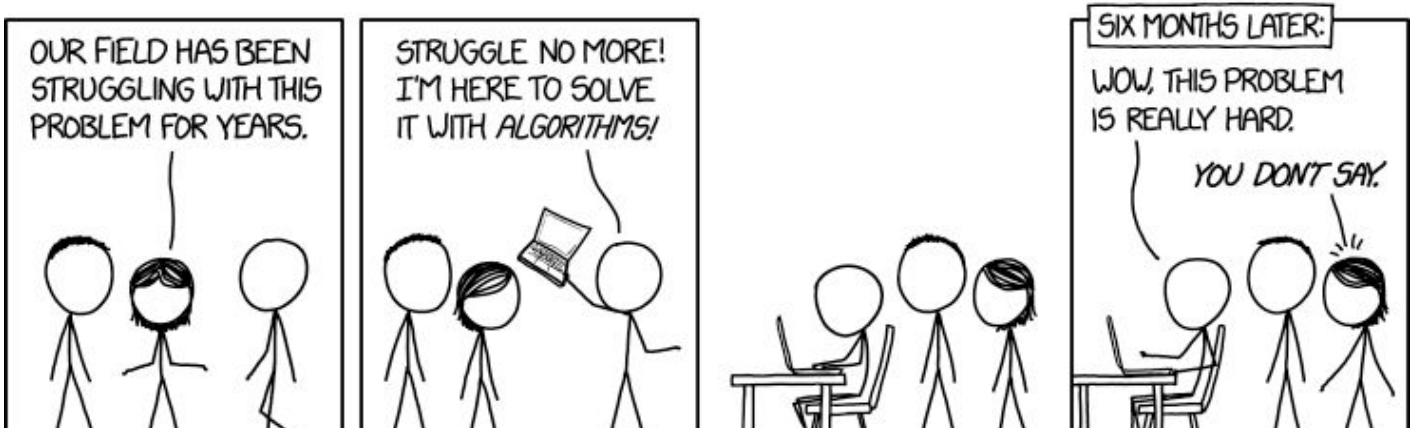
$$\int_{r_e}^{\infty} \frac{1}{2}\left(\frac{q}{4\pi r^2}\right)^2 4\pi r^2 \, dr = \frac{q^2}{8\pi r_e}$$

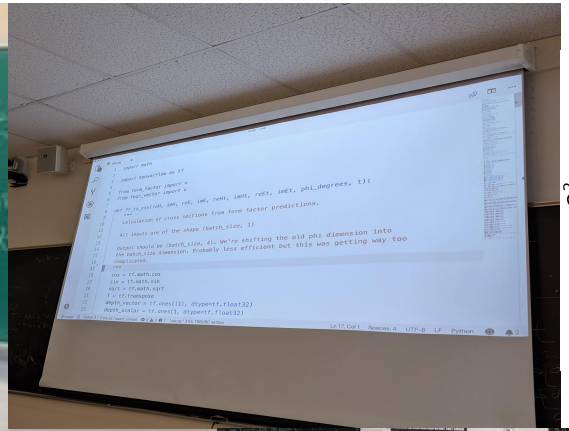
$$\min_f \sum_{i=1}^{n} V(f(x_i), y_i) + \lambda R(f)$$

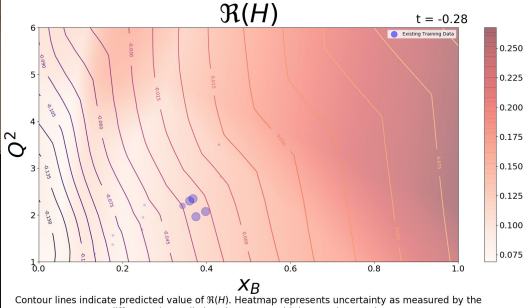


Defining a machine learning problem

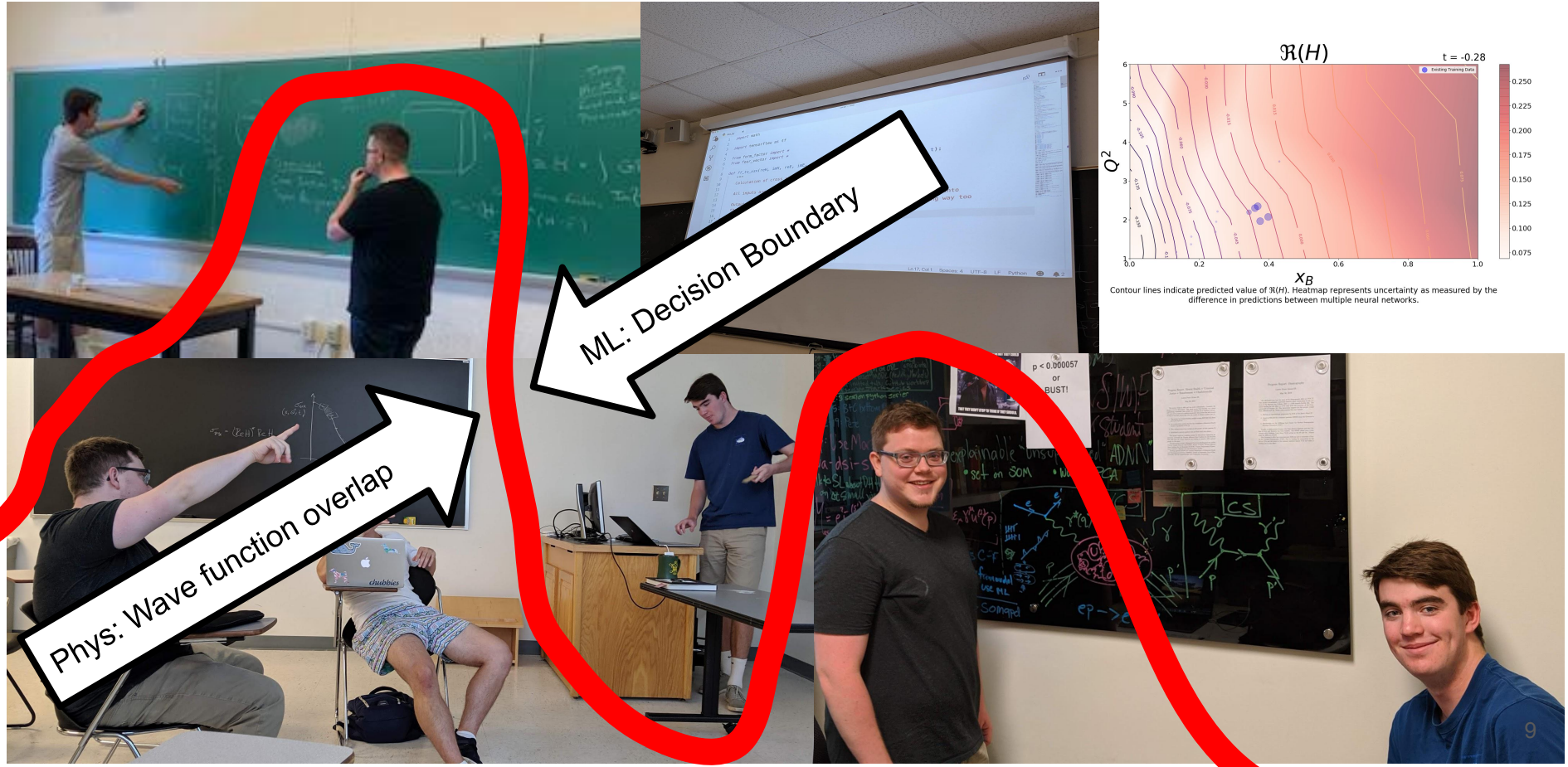
1. Physicists provide the data
   a. Data scientists do not know how to propagate errors (still ad hoc)
2. Data scientists apply a machine learning model
   a. Physicists do not know the capabilities of ml tools (selection is hard)
3. Physicists define the loss function
   a. Data scientists struggle to get the algorithms to converge (tuning and training takes time)
4. Data scientists provide a trained model to predict physics analysis goals
   a. The scientific community does not know how to explain ml models (for cutting edge ml)





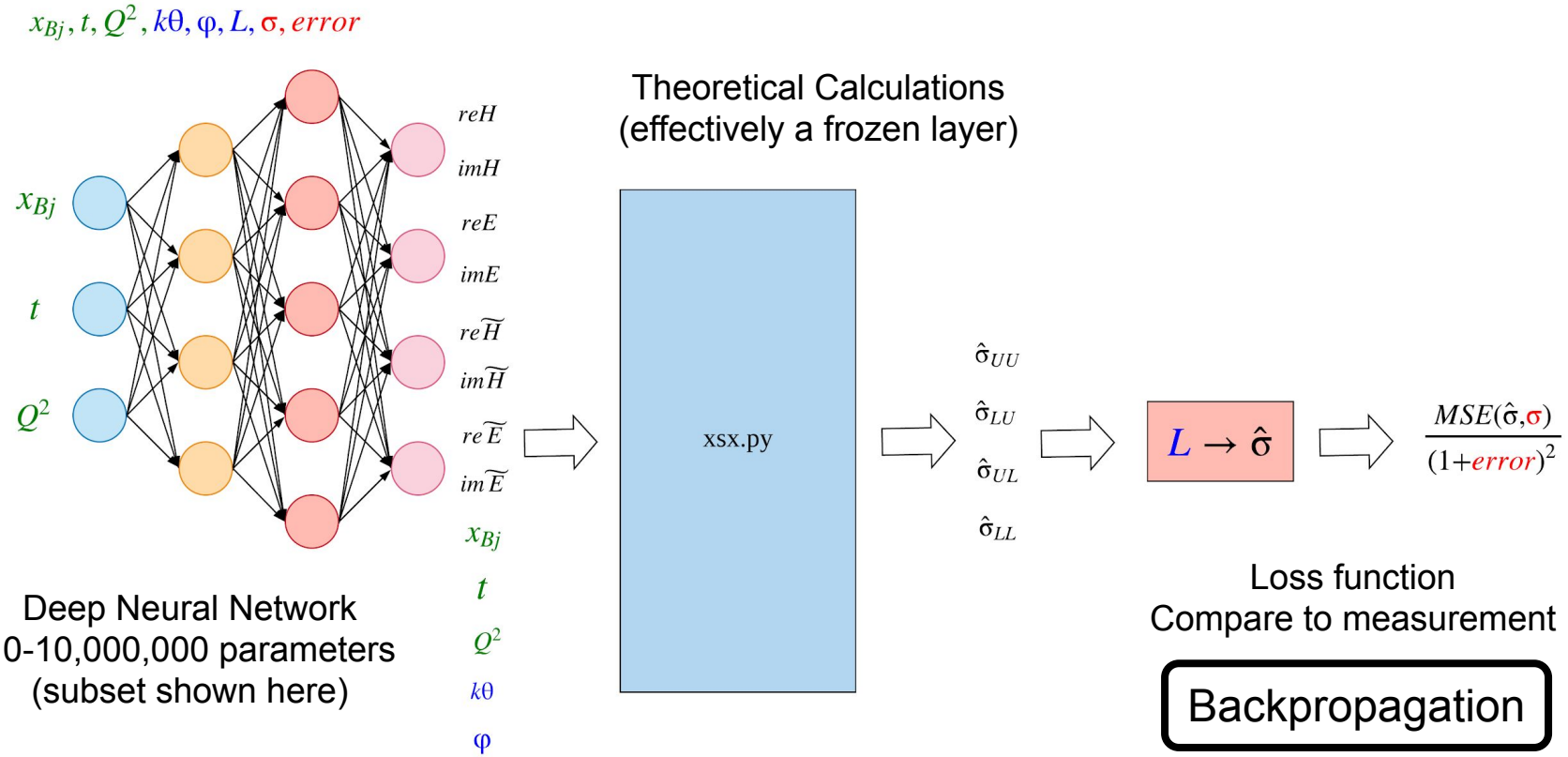


Contour lines indicate predicted value of $\Re(H)$. Heatmap represents uncertainty as measured by the difference in predictions between multiple neural networks.



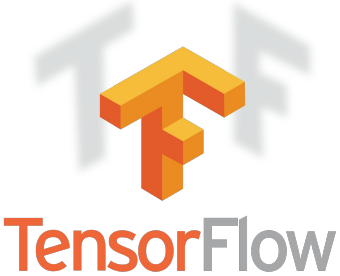
Contour lines indicate predicted value of $\Re(H)$. Heatmap represents uncertainty as measured by the difference in predictions between multiple neural networks.

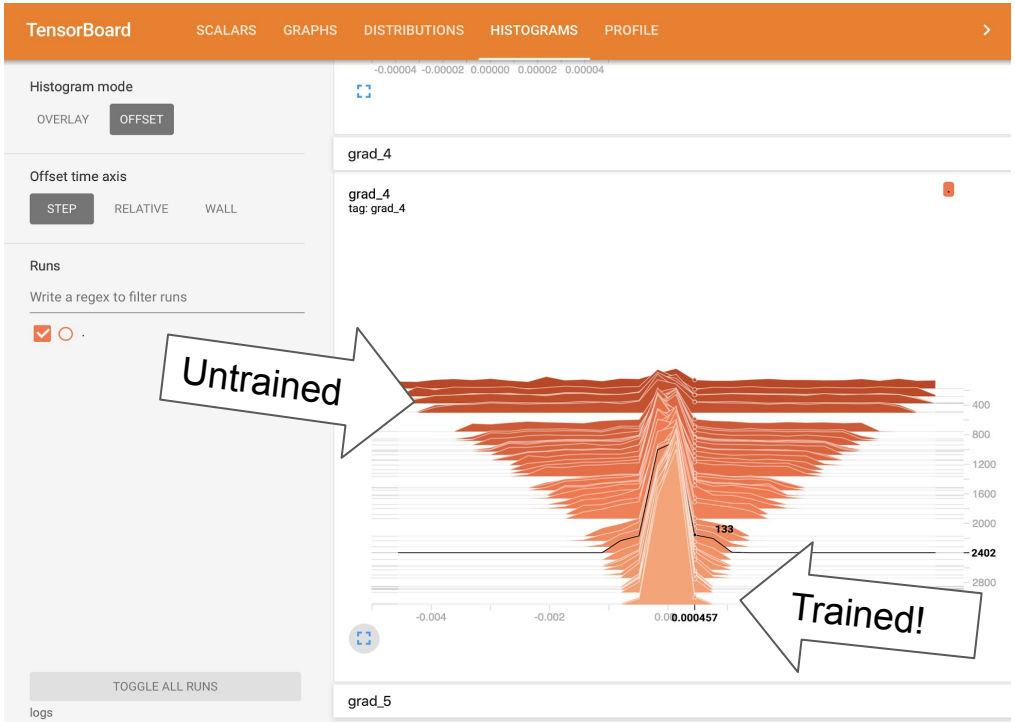
Machine learning solution - Supervised Learning



Implemented in Tensorflow

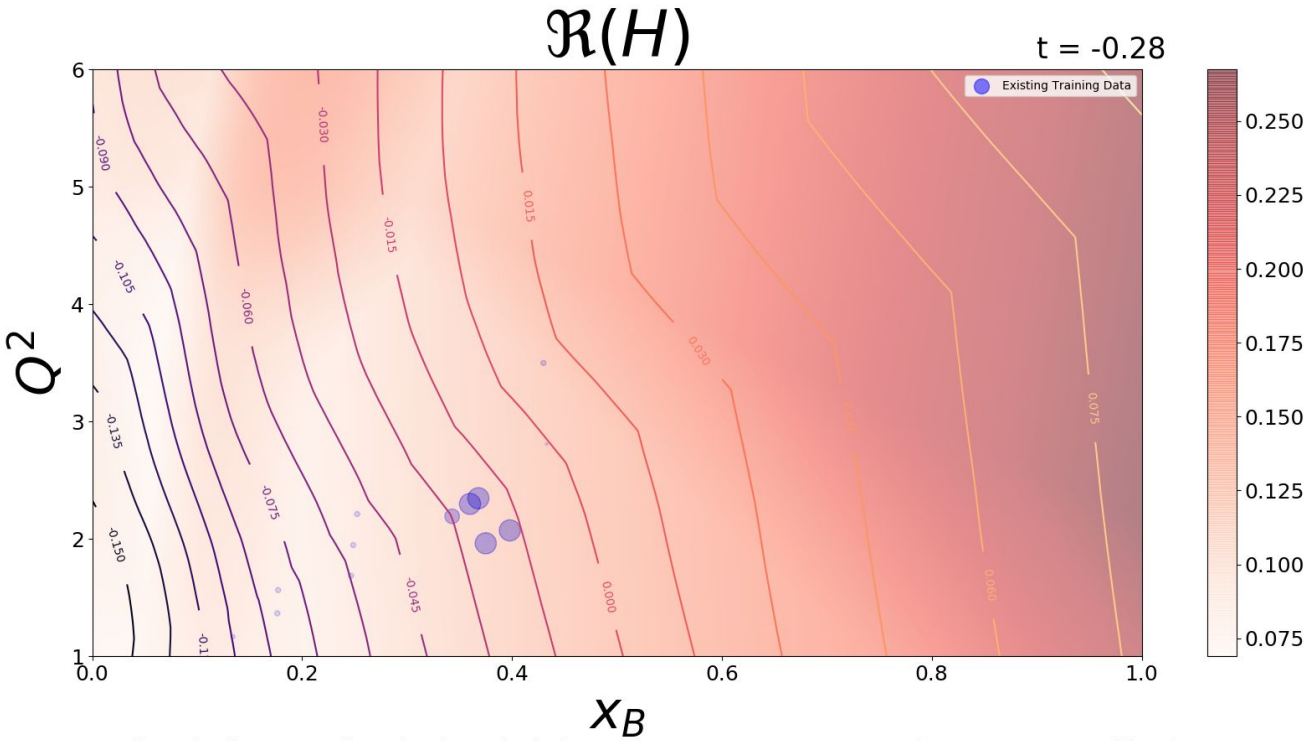
- Cutting-edge deep learning framework
- Automatic differentiation
- Runs on GPUs for amazing speed
- Large Open-source community





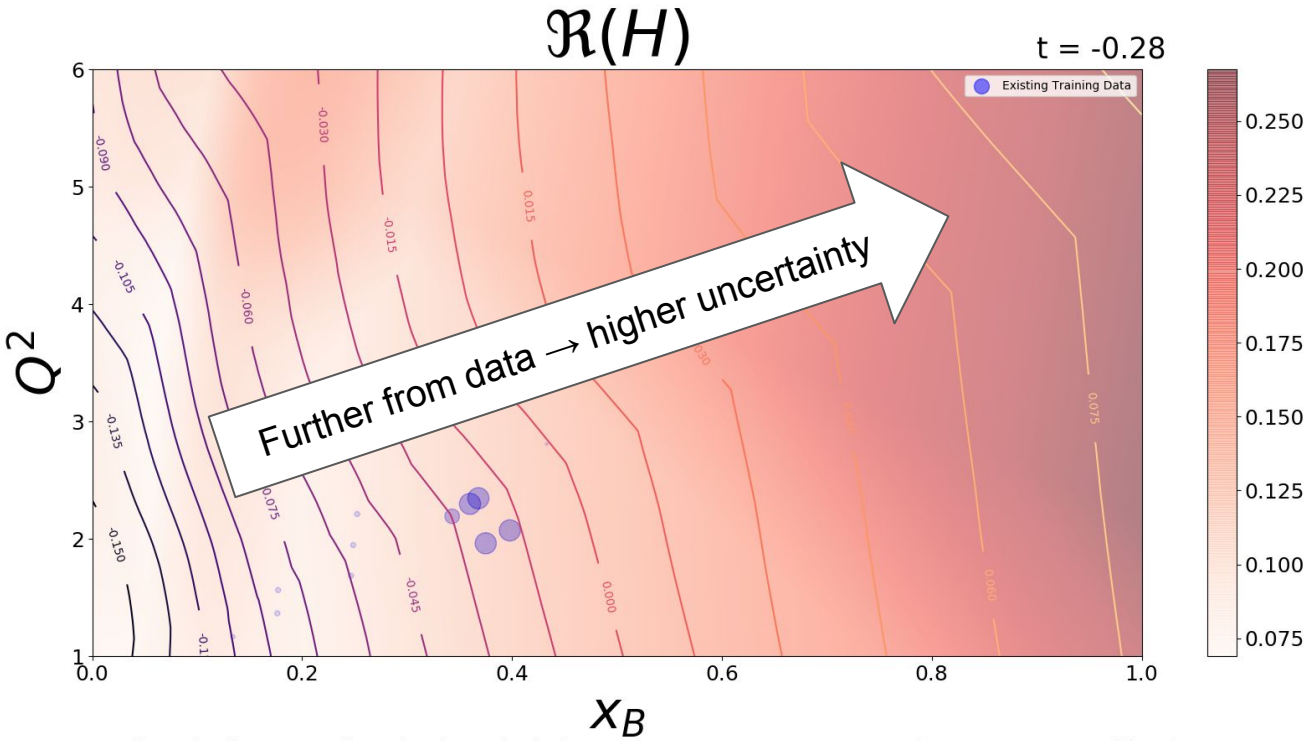
Training is like collapsing the wave function.

Result of ML pipeline - (tuning and training underway)



Contour lines indicate predicted value of $\Re(H)$. Heatmap represents uncertainty as measured by the difference in predictions between multiple neural networks.

Result of ML pipeline - (tuning and training underway)



Contour lines indicate predicted value of $\Re(H)$. Heatmap represents uncertainty as measured by the difference in predictions between multiple neural networks.

Up Next: Unsupervised machine learning

Uniform Manifold Approximation and Projection (UMAP)

- Cluster GPDs into lower dimensions (2D/3D)
- Manifold learning algorithm
- Akin to: t-SNE and PCA (dimensionality reduction)
- Enable visualization of high dimension data

Summary

https://pages.shanti.virginia.edu/Femtography/

- Students now proficient in interdisciplinary collaboration
- Built an integrated pipeline with Tensorflow to predict CFFs and cross sections

- Unsupervised dimensionality reduction algorithm is in development
- After publication will release software as an Open-source project